Duration of a Sung Token in Pitch Matching as a Function of the Duration of the Response

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation with distinction in Speech and Hearing Science in the undergraduate colleges of the Ohio State University

by

Stephanie Finical

The Ohio State University

June, 2007

Project Advisor: Dr. Michael Trudeau, Department of Speech and Hearing Science

Abstract

Trained singers are often required to produce a note that matches the pitch of another sound, such as a musical instrument. Previous studies have uncovered a number of stimulus aspects that influence the ability to match pitch, including the spectral complexity of the sound and its duration. However, one unanswered question is whether, when matching pitch, trained singers also attempt to match other aspects of the stimulus, or whether they can selectively attend to pitch. The present study investigated whether the duration of the eliciting stimulus affected the duration of a trained singer's pitch matching response. The samples for the study were obtained from a previous matching study by Ives (2002). Ten male trained singers were presented with pure tone and synthesized voice stimuli at 8 fundamental frequencies, and 4 durations. The responses were ten analyzed using Kay Elemetrics Computerized Speech Lab for both fundamental frequency and duration. Results indicated a slight trend toward increasing response duration as stimulus duration increase, but this trend was not statistically significant. This finding suggests that trained singers can indeed focus on producing a matching fundamental frequency without matching the stimulus duration, and implies that the two stimulus aspects are processed independently.

Acknowledgements

First I would like to thank Dr. Mike Trudeau for all the help, time, and guidance he provided throughout this research experience. Secondly, I would like to thank Dr. Janet Weisenberger for her direction and assistance with this project. Thirdly, I would like to thank Stephanie Joseph and Allison Masty for making the experience of data gathering and analysis a good time. Fourthly, I would like to thank my fantastic and wonderful friends and family for all their support and help in getting me through this experience without going crazy. Fifthly, I would like to thank the College of the Arts and Sciences for the financial support they provided for my research.

Table of Contents

Chapter One: Introduction

How do people match the pitch of another sound? How does a singer match the pitch of his voice to that of another voice, musical instrument or a pitch pipe? Pitch matching accuracy (PMA) can be defined as the ability to reproduce a pitch without being flat or sharp. PMA is a necessary skill for a singer who hopes to achieve any success in his or her vocal occupation. However, relatively little is known about factors that affect our ability to pitch match.

What is known is that pitch matching is a two part process that involves first correctly perceiving the pitch and then accurately producing it. One of the factors that might affect production is prior training, i.e. the benefit of the training and practice of singers versus the inexperience of non-singers. A second factor is the sound used to elicit the pitch match. Aspects of the eliciting sound itself that have been examined include the spectral complexity of the sound and the duration with which it is presented. The results of these studies have provided sources of valuable information but certainly have not presented us with a full understanding of the mechanisms involved in pitch matching.

Factors influencing pitch matching ability

A study by Murry (1989) focused on what qualities of the singer affect the production of a pitch. Murry noted that the most important aspect in matching pitch is matching fundamental frequency. In western harmonics each perceived pitch corresponds to an absolute acoustic frequency (Ives, 2002).The first question in Murry's study was whether or not trained singers were better able to match pitch. To test this, Murry used two groups in the study, one made up of singers, the other of non-singers. The second question of the study was whether or not practice over trials increased accuracy.

During each of the trials each subject was presented with one of three possible frequencies, all of equal duration. The stimulus was presented as one tone every six seconds. Murry's results showed that singers possessed greater laryngeal accuracy in PMA tasks than non singers. Accuracy was measured by comparing the target frequency with the first identifiable waveform from the response as well as the mean of the first five identifiable waveforms. Singers also showed a trend towards increased accuracy over trials; however; it was not statistically significant. Non-singers did not demonstrate any set trend towards improvement and there were no statistically significant changes noted as a function of practice. The overall results support the assumption that in all likelihood the training and experience singers have allows them to outperform non-singers in pitch matching tasks.

However, the stronger performance of singers in Murry's study did not differentiate between training and talent as the determinant of performance. Watts (2004) conducted a study investigating the differences between two other categories of singers, untrained accurate and untrained inaccurate singers in both pitch matching and pitch discrimination tasks. The accuracy of the untrained singers is considered to be related to the concept of talent. Watts (2003) had previously defined talent as a special natural ability that has the potential to lead to a capacity for achievement or success. In the 2004 study the relationship of talent in pitch perception to talent in pitch production was evaluated. In the voice production section of the tasks participants were asked to sing the first stanza of "America the Beautiful" as well as to perform a pitch-matching task. The eight stimulus tones they were asked to match were synthesized piano notes ranging in frequency from 220Hz to 440Hz. In the pitch discrimination task participants were presented with pairs of tones and asked to determine whether or not they were the same pitch. The stimulus tones were either synthesized piano or synthesized trumpet notes. The pairs were presented in the following combinations: same pitch and same timbre, same pitch and different timbre, different pitch and same timbre, or different pitch and different timbre.

Results showed that the untrained accurate singers were found to be the most accurate group in both the pitch production and pitch discrimination tasks. The greater accuracy in the pitch matching tasks supports the argument that talent, even in the absence of training, affects both the perceptual and production aspects of pitch matching.

Stimulus determinants of voice pitch perception

Von Bekesy's work in the 1950's and 1960's brought to light some of the basics of pitch perception. His investigations revealed the number of cycles of stimulus that a person must hear in order to extract the pitch. For a stimulus with a frequency of less than 1000Hz the tone must be presented long enough for the wave to go through 3-9 cycles in order for the listener to get an accurate estimate of the pitch. For a stimulus with a frequency 1000Hz or greater it must be presented for at least 10ms with no fixed number of cycles (von Bekesy, 1960).

Moore et al. (1992) investigated how the harmonic structure of an eliciting stimulus affected a subject's ability to discriminate pitch. Subjects were asked to adjust the pitch of a complex tone to match the pitch of a second tone. Harmonic structure of the second tone was either the same as that of the first tone or different than that of the first. The pitch matches were best and least varied when the two stimuli had the same harmonic structure, which suggests that harmonic content does affect the ability to match pitch.

Tervaniemi (2000) examined pitch perception in more detail. More specifically, he examined the individual's ability to discriminate between pitches. The two aspects of the stimulus that Tervaniemi focused on were the pitch's spectral complexity and the duration of the sound. To determine the discrimination abilities of the subjects, Tervaniemi examined the neural functioning of the mismatch negativity (MMN), a component of the event related potential. The MMN is produced when the listener is able to discriminate any change in sound. During the study subjects watched a silenced movie while they were presented with various sound stimuli. They were instructed to ignore the sounds presented.

The sounds the subjects were presented with included standard tones (pure tones) as well as two deviants (spectrally rich tones) for each standard tone. The deviants varied from the original tone by either +2.5% or -2.5%. The deviant tones also varied by the number of partials per tone, with some having one partial, some three and some five.

Tervaniemi's results showed that the MMN amplitude was enhanced during the presentation of tones with three or five partials, the more spectrally complex sounds. The increases in amplitude of the MMN were considered statistically significant. However, increased sound duration beyond 100ms did not result in any increases in MMN amplitude. This led Tervaniemi to conclude that increased spectral complexity, not increased temporal information, affects MMN and this likely then reflects pitch discrimination.

Neuhoff et al. (2002) investigated the problem of the confounding aspect of the interaction between pitch and loudness on a subject's perception of auditory stimuli. In the study, subjects were presented with stimuli that had changes in both pitch and loudness that went in either the same or opposite directions. The subjects were asked to pay attention to the changes in each factor and were asked to report what changes occurred. Results showed that when the changes of pitch and loudness occurred in the same direction the perceived change in either stimulus aspect was greater than when then changes occurred in opposite directions. These findings indicate an inability to fully distinguish the different factors of a stimulus when perceiving a sound.

Stimulus effects on pitch production

A study by Ives (2002) investigated the influence of both of the stimulus factors in Tervaniemi's study (duration and spectral complexity) on the singer's ability to match fundamental frequency. In his study Ives asked ten male trained singers to perform a specific pitch matching task. The frequencies the singers were presented with varied in spectral complexity (pure tone vs. synthesized voice), stimulus frequency (130.1Hz to 311.1Hz), and stimulus duration (50ms, 100ms, 200ms, 300ms). Each subject was given five lists of sixty-four tokens each to match. In each list there was equal representation of all the possible token variations.

Ives than analyzed the data with *Cool Edit 2000*. According to his results, PMA was not affected by stimulus type or duration. However, he did discover that stimulus duration had a statistically significant effect on the response latency (RL) of the singers. As duration was increased, RL decreased. The mean RL was significantly shorter for token durations of 300ms(0.6375s) and 200ms(0.771s) than for durations of 50ms(0.8712s). This implied that exposure to at least 150ms of the pitch beyond the initial 50ms produced a significant reduction in RL. Ives also compared PMA after the first identifiable period (P1) and after the first five identifiable periods(P5). The PMA at P5 was statistically significantly more accurate than that of P1. These results suggest that there is a period of adjustment needed for the singers as they produce the matching pitch to reach phonatory stability.

Ameer (2003), in a follow up study to Ives, investigated the time necessary to reach phonatory stability and the effects of token type on the stabilization time. Ameer found that a mean time of .0261 was required for a singer to reach phonatory stability in a sung response. This time period was equivalent to a mean of six cycles of response by the singer. Ameer also found that the stabilizing time was unaffected by token type. These findings support Ives' finding of greater accuracy at five cycles and suggest that if the response were measure over a longer period it would produce even greater accuracy.

Given this finding of greater accuracy as more periods of the response were measured, Curran's (2004) follow up study to Ives examined the overall accuracy of the singer's response. It looked at the best measurement interval to represent PMA (P1, P5, entire response), the effects of token duration on PMA, and the effect of token type on PMA. Curran used the initial responses recorded from the Ives study.

Curran found that using P1 or P5 produced significantly less accurate measurements of PMA than when measuring the entire response. An initial look at the results seemed to show that neither token type (synthesized voice or pure tone) produced a significant advantage in the pitch matching task. However, examination of the interaction between token type and token duration showed that there was a statistically significant, albeit discreet, effect. The results of the examination showed that responses to the pure tone 50ms tokens were significantly less accurate than responses to the pure tone 300ms tokens, and than tokens of all durations of the synthesized voice stimuli. This led Curran to the conclusion that there is an advantage to the synthesized voice stimuli, but it is only present at the briefest durations. Her findings also led to the conclusion that while 50ms duration is sufficient for the human ear to discern pitch there is an advantage for pitch matching when stimulus is presented in longer durations of 200ms or 300ms.

Because Ives, Ameer, and Curran all used *Cool Edit 2000* they all faced similar challenges with the software. Cool Edit is a digital program but it is considered to be relatively unsophisticated. Due to the lack of sophistication working with this software was a labor intensive endeavor. To start, Ives, Ameer, and Curran had to amplify the signal in order to first view it and then to analyze it. They went through the initial process of identification, and analysis of the signal. Then the signal had to be reviewed using prior knowledge of the eliciting stimulus parameters to determine the expected parameters of the response. Many times there was too much information for the software to analyze and the resulting frequency would be an octave too high. In these cases Ives, Ameer, and Curran would have to re-analyze the signal with more specific restrictions to get an accurate measure. These issues raise questions about the validity of software to properly analyze stimuli as well as the interaction between the user and the software.

There are also still many questions to be answered concerning the processes of the singer in pitch matching tasks. What stimulus features are the most solvent to the observer in a pitch matching task? Are there aspects besides frequency that a singer attempts to match? What effects do aspects such as duration or token type have on the response produced by the singer in a pitch matching task?

As reported by Nuehoff et al. (2002) above, subjects have some difficulty in separating th perceptual attributes of pitch and loudness. It is of interest to determine whether individuals can separate perceptual aspects of the stimulus in production tasks. In attempting to match pitch, is the individual able to attend selectively to the pitch of thee eliciting stimulus, or is the stimulus matched on all is perceptual dimensions? Given the findings of Curran (2004) suggesting that token duration may affect pitch matching accuracy, are subjects also attending to stimulus duration in their pitch-matching productions?

Research Question

The present study further addressed the issue of token duration on a singer's pitch matching response. More specifically, this study investigated the effect of the duration of the stimulus on the duration of the response produced by the singer. Data from Ives (2002) were analyzed for 8 trained male singers, who were presented with two types of tokens (pure tone and synthesized voice), 8 different stimulus fundamental frequencies and four durations. Results were analyzed to determine which, if any, of these stimulus factors had an impact on the duration of the singer's response.

Chapter 2: Methodology- Ives Study

Subjects

The volunteer participants for Ives' study were ten male students in Vocal music, all participating in Men's Glee Club. Each student's vocal background included at least four years of formal instruction and six years of choral singing. None of the participants reported any history of any laryngeal pathology, voice disorders requiring surgery or any voice therapy. Also, audiometric hearing screenings ensured all participants were within normal thresholds at 20dBHL for all test frequencies used. At least two hours prior to the experimental task, all singers had at least ten minutes of purposeful vocal warm up. Subjects were made aware of the task and requirements to complete it.

Tokens

Tokens used in the task consisted of two types, pure tone or synthesized voice. The synthesized human male voice tokens were created with the Sensimetrics, high level parameter speech synthesis system (version 2.2). The first three formants of the synthesized voice were 554 Hz, 916Hz, and 2466Hz and were based on a sample from The Ohio State University music faculty. Vibrato was also added to the synthesized tones at a rate of ten cycles per second. Pure tones were generated by using Cool Edit 2000. The tokens consisted of eight pitches, (C,D,E,F#,G#,A#,C#,D#) with a frequency range of 130.8Hz-311.1Hz. All tokens were also presented at four different durations: 50ms, 100ms, 200ms, and 300ms.

The tokens were presented to the to the participants in five randomly sequenced sets. Each set equally represented all variations of token type, duration and frequency. The order in which the five sets of tokens were presented to the participants was also randomized.

Recording Procedure

Participants were placed in a sound treated room and instructed to respond to the presented stimulus as quickly as possible. A time constraint was imposed by presenting a five second delay between stimuli. In order to simulate more closely the study done by Murry (1989), subjects were instructed to use the /i/ vowel for their response. There were no restrictions placed on the singers concerning the use of “straight tone” or “vibrato” or on the loudness or softness of their voices.

Analysis Procedure

Cool Edit 2000 was used to record, store, and analyze Ives’ data, including both stimuli and responses. The inverse of temporal measurement was used to determine F0 of the first identifiable period (P1) and the first five periods (P5). Response latency was measured as the time lapse between the end of the stimulus and the beginning of the response. Temporal measurements were made to ms accuracy using the zoom capabilities of Cool Edit. PMA was measured in two ways; first, by comparing the stimulus F0 with that of the first identifiable period of each singer’s response. The second method for accuracy measurement was difference between the target, the first identifiable period (P1), and the first five identifiable periods (P5).

All data measurements were taken from tokens set at the midpoint of the frequency range at G#- 207.65Hz. The target was chosen due to its nearness to the midrange, which might give it greater perceptual salience. This provided a total of forty points per participant. Each of the two F0 values taken from each measurement (P1, P5) was converted to semi tones. Re-analysis was conducted on one randomly selected token from each of the (five) response files. Re-analysis included identifying the end of the stimulus token, initiation of the first identifiable period, and terminations for the first and fifth identifiable periods. This yielded the production of fifty measurements for each variable.

Methodology-Present Study

Analysis

In the present study data recorded by Ives were analyzed using the Kay Elemetrics Computerized Speech Lab (CSL). Two windows were used in the speech lab during analysis. In one window Ives' data were opened in the form of wave files. Voice Period Marks were selected under the analysis option. Impulse location for the marks was set to mark at zero crossing. Then analysis range for voice period marks was set for a minimum of 70Hz and a maximum of 500Hz. Next (also under the analysis option), Pitch Contour was selected. Analysis range for pitch contour was also 70-500Hz. Framing length and framing advance for measuring pitch contour were both set at five milliseconds and display range was set to show a draw dot contour from 0-500Hz. One of the issues that came up during data analysis was related to the amplitude of the tokens being measured. Because CSL had difficulty in analyzing low amplitude tokens the framing was adjusted to 25ms frame length and 20ms frame advance in order to obtain data on the sample. After analysis parameters were set individual tokens were isolated using the cursors in the top window. Once a token had been selected pitch contour analysis was run and the contour was displayed in the bottom window. The statistics button was then selected and CSL displayed start and end times of the selected token as well as the mean fundamental frequency. There were forty total token sets analyzed by three researchers. For reliability each researcher analyzed two-thirds of the data so that each token set was analyzed a total of two times by two individuals.

Chapter 3: Results

Figure 1 shows the variation in response duration as a function of the two types of stimulus tokens (synthesized voice and pure tone). There was no significant effect of token type on the duration of the subjects' responses.

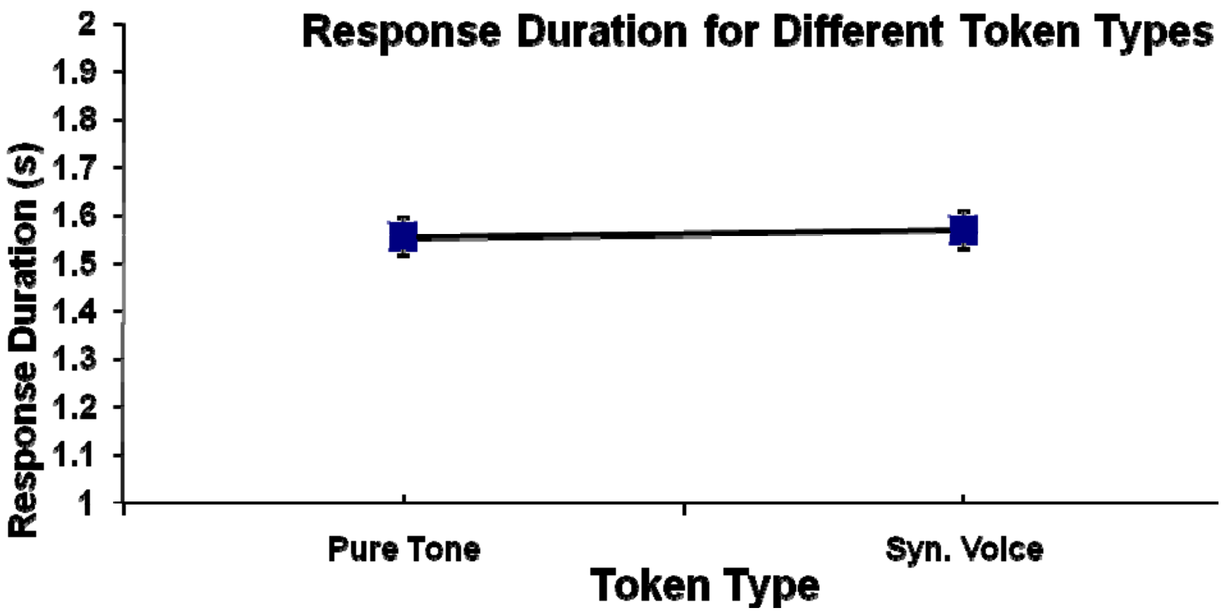

Figure 1: Response Duration as a function of Different Token Type (pure tone vs. synthesized voice).

The effect of fundamental frequency on response duration is shown in Figure 2. As with token type, fundamental frequency does not appear to have any significant effect on the duration of response.

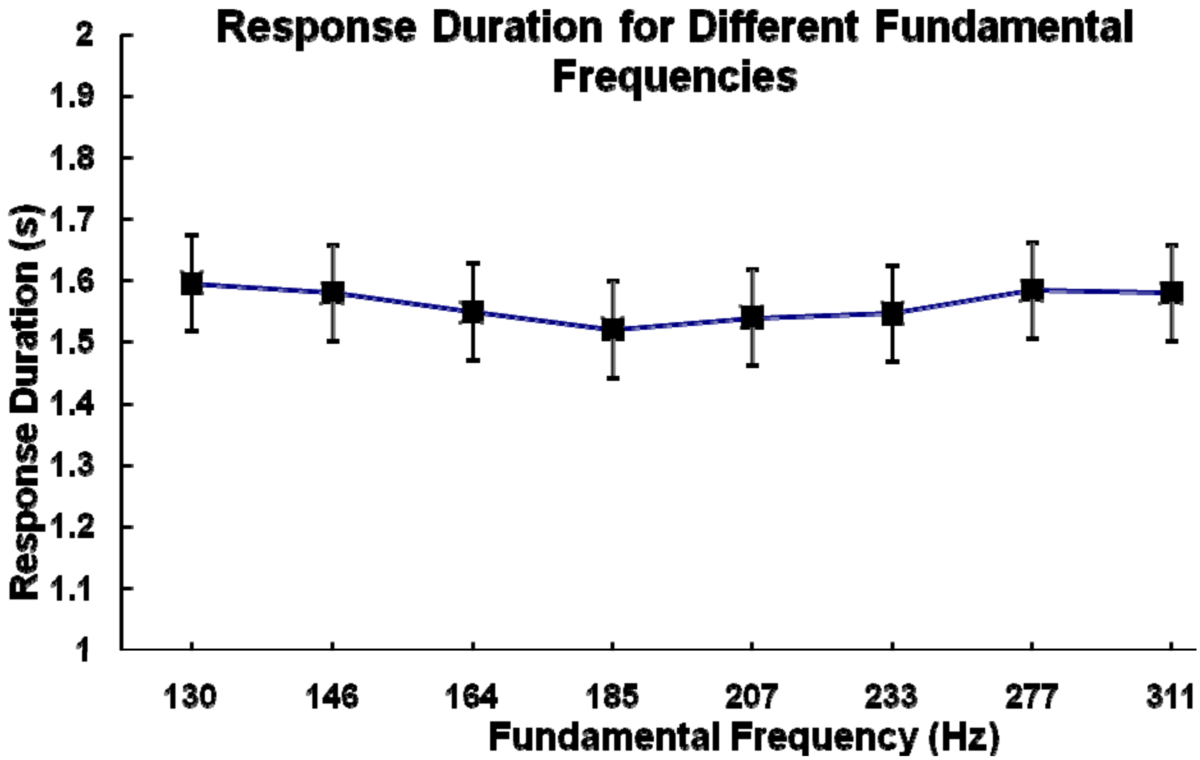

Figure 2: Response Duration as a function of Fundamental Frequency

Figure 3 shows the variation in response duration as a function of stimulus duration. In the case of this stimulus factor there was a slight trend toward increased response duration with increased stimulus duration.

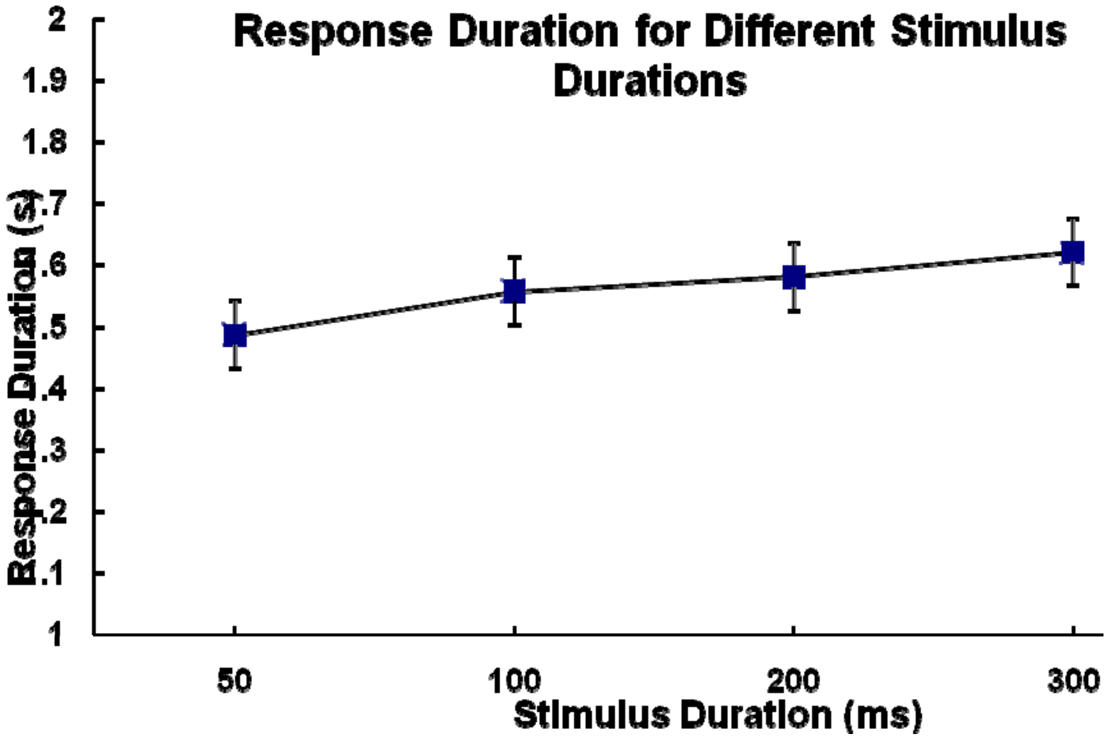

Figure 3: Response duration as a function of Stimulus Duration.

A three factor analysis of variance was performed to evaluate the significant effects of differences in stimulus data. As is evident from the figures this analysis showed that the stimulus variables had no significant effects on the response duration. Further, no interaction effects of the variables were found.

Chapter 4: Discussion

The purpose of this study was to investigate the effect of stimulus duration, as well as other stimulus factors, on the duration of a response in a pitch matching task. In pitch matching tasks the question arises of the similarities across responses given by a singer. Are they consistent on all aspects or is their much variation from response to response? It is assumed that if there are inaccuracies that they would occur at the beginning of the responses. As the singer progressed through the response the singer would adjust the pitch production. If there are inaccuracies as well as different durations of responses those responses with shorter durations would be the most affected. In the present study results showed that none of the stimulus factors analyzed had any effect on the response duration.

The lack of effects indicates that the singers were able to match only the pitch of the stimulus without simultaneously attempting to match duration. All the singers had extensive vocal training and experience, at least four years of formal instruction and six years of choral singing. It is possible that this extensive training enabled the singers to selectively match pitch without attempting to match other characteristics of the stimulus. Further research investigate the same task with non singers to specifically determine what role training played in the obtained results.

Another reason for a lack of effects on duration could be explained by Ameer's (2003) results. According to Ameer's findings the singers need at least six cycles of response to stabilize. The required amount of time to reach the six cycles and reach stability sets a lower time limit for any response by the singers. In the present study, six cycles for a 207 Hz stimulus would place a lower limit of 25ms on a response. As can be seen, all responses exceeded this value.

Another possible cause of consistency in response duration could have also been the five second time limit imposed upon the singers as part of the task. The singers may have produced responses rhythmically in order to be prepared for the next stimulus tone. There are several ways that further studies could address this issue. These include a longer time window, a varying time window, and allowing the subjects to set their own pace of stimulus, presentation, and response production. A study could also employ multiple time window options in order to compare the different effects that may be caused.

Additional issues were raised in this study concerning the accuracy of Kay Elemetrics CSL. When analyzing the response tokens in the data the judges in the study, unlike those in the studies by Ives (2002) and Murry (1989), were blind to the parameters of eliciting stimuli. When planning the methodology it was assumed that given the more sophisticated software being used that CSL should be able to extract and determine parameters of the response token, such as fundamental frequency and duration, without the input of additional information about the stimulus factors. However, it was discovered that additional restrictions did need to be set in CSL to get consistently accurate results. This was likely to due to the weakness of the signal being analyzed. It can be speculated that if the judges were made aware of the parameters of each eliciting stimulus and were able to focus the filters and range of the software the results would be far more accurate.

Overall the study showed that singers were able to be selective about which factors in the stimulus the attended to during a pitch matching task. These results have left several unanswered questions about the processes of pitch matching as well as the methodology used to measure pitch matching accuracy. Hopefully, the questions brought up by this study will lead to future investigations that will uncover more information about the underlying mechanisms of pitch matching.

References

Ameer J. Time to Phonatory Stability in Trained Singers Cued for Pitch Matching by Pure-Tone and Synthesized Voice. Unpublished Honors Thesis, (2003) The Ohio State University

Curran K. Measurement of Pitch Matching Accuracy: How much is too little? Unpublished Honors Thesis (2004) The Ohio State University

Ives S. The Effects of Sound Duration and Spectral Complexity on Pitch-Matching Accuracy in Trained Singers When Cued with Pure-Tone and Synthesized Human Voice Models. Unpublished Master's Thesis (2002) The Ohio State University

Moore B, Glasberg B, Proctor G. (1992) Accuracy of Pitch Matching for Pure Tones and for Complex Tones with Overlapping or Nonoverlapping Harmonics. *J. Acoust. Soc. Am.* 91 (6): 3443-3450.

Murry T. (1990) Pitch Matching Accuracy in Singers and Nonsingers. *J Voice* 4 (4): 317-321
Neuhoff J, Kramer G, Wayand J. (2002) Pitch and Loudness Interact in Auditory Displays: Can Data Get Lost in the Map? Journal of Experimental Psychology. 8 (1): 17-25.

Tervaniemi M, Schröger E, Saher M, Näätänen R. (2000) Effects of Spectral Complexity and Sound Duration on Automatic Complext-Sound Pitch Processing in Humans- A Mismatch Negativity study. *Neuroscience Letters* 290: 66-70.

Von Békésy G. (1960) Experiments in Hearing. McGraw Hill Book Company, Inc. New York, USA.

Watts C, Murphy J, Barnes-Burroughs K. (2003) Pitch Matching Accuracy of Trained Singers, Untrained Subjects with Talented Singing Voices, and Untrained Subjects with Nontalented Singing Voices in Conditions of Varying Feedback. *J Voice*. 17: 185-193.

Watts C, Moore R, McCaghren K. (2005) The Relationship Between Vocal Pitch-Matching Skills and Pitch Discrimination Skills in Untrained Accurate and Inaccurate Singers. *J Voice*. 19 (4): 534-543.